



Universidade do Minho
Escola de Engenharia

GroceryHelper

Mineração de Dados

Especialização de Engenharia de Conhecimento

Joana Catarina Oliveira Gomes, Pedro Marcelo Bogas Oliveira,
Tomás Cardoso Francisco

Mestrado em Engenharia Informática, Universidade do Minho, Braga, Portugal.

Contributing authors: pg52685@alunos.uminho.pt; pg54144@alunos.uminho.pt;
pg54263@alunos.uminho.pt;

1 Introdução

A tecnologia está cada vez mais integrada nos processos de negócio, oferecendo soluções avançadas para a otimização e automação de tarefas complexas. No setor de venda de produtos alimentares, a seleção eficiente de produtos é crucial para atender às demandas dos consumidores e manter a competitividade no mercado. Este relatório apresenta um sistema desenvolvido para auxiliar nesse processo, utilizando técnicas de processamento de linguagem natural (PLN) e *scraping* de dados para obter informações detalhadas sobre produtos alimentares e bebidas.

A criação deste sistema foi motivada pela necessidade de lidar com a complexidade crescente na seleção de produtos alimentares, onde a variedade e a quantidade de itens disponíveis dificultam a tomada de decisão informada pelos vendedores. O sistema desenvolvido não apenas extrai informações detalhadas dos produtos, mas também facilita a geração de respostas automatizadas a perguntas específicas sobre esses produtos, contribuindo para uma experiência mais eficiente e informada para o utilizador final.

2 Estado da Arte/Trabalho Relacionado

Existe uma variedade de sistemas e tecnologias que empregam técnicas similares de PLN e *scraping* para extração de dados. Exemplos incluem sistemas de recomendação de produtos baseados em análise de sentimentos e preferências dos consumidores, assim como ferramentas de monitorização de preços e disponibilidade de produtos em tempo real. A integração de PLN e *scraping* oferece vantagens significativas na precisão e na eficiência da colheita de dados, sendo amplamente adotada em aplicações de análise de mercado e competitividade empresarial.

É importante destacar que, embora o sistema desenvolvido seja mais prático e funcional do que intuitivo e bem projetado em termos de interface de utilizador, este atende aos requisitos essenciais de fornecer informações detalhadas e precisas sobre produtos selecionados.

3 Métodos/Desenvolvimento

O sistema foi implementado utilizando a linguagem de programação Python e o framework Flask para desenvolvimento web. A integração com a API da OpenAI permitiu a utilização de modelos avançados de PLN para compreensão e geração de respostas automáticas baseadas em perguntas sobre produtos. Além disso, o framework Scrapy foi empregado para realizar *scraping* de dados de websites como o Auchan, obtendo informações detalhadas como nome, preço e link dos produtos.

É importante observar que, devido a recentes atualizações nas páginas HTML e políticas do site da Amazon, o sistema já não podia mais ser utilizado no site após essas mudanças. No entanto, existem, em vídeo, provas que demonstram o funcionamento do programa previamente a essas atualizações, o que valida a eficácia do sistema até aquele ponto.

4 Resultados e Discussão

Durante os testes do sistema, foram observados resultados positivos quanto à precisão na extração de dados e na eficiência na resposta a perguntas sobre os produtos selecionados. A integração com técnicas de *scraping* permitiu obter informações detalhadas dos produtos, enquanto o uso de modelos de PLN facilitou a geração de respostas contextuais e informativas.

O feedback dos utilizadores destacou a utilidade do sistema em agilizar o processo de seleção de produtos e fornecer informações relevantes para tomada de decisões. As limitações encontradas, como a dependência de estruturas estáveis nos websites de comércio, realçam a necessidade contínua de adaptação e atualização do sistema para garantir a sua eficácia a longo prazo.

4.1 Página inicial

Grocery Helper

Eu quero pesquisar sobre pão, peixe e queijo.

Submit

Fig. 1 Página inicial

A página *home* serve como ponto inicial da interação com o utilizador. Exploremos o que está incluído na página, para que serve e como funciona:

4.1.1 Conteúdo da Página

A página apresenta o título "Grocery Helper" e contém um campo de entrada de texto onde o utilizador pode escrever o que deseja e um botão de envio.

4.1.2 Objetivo Geral

A página inicial ('/') permite que os utilizadores insiram um texto relacionado a produtos alimentares e bebidas. Esse texto será processado para extrair informações sobre os produtos mencionados.

4.1.3 Funcionamento

Quando um utilizador insere um texto no campo de entrada e clica em "Submit", os dados são enviados para a rota '/products' via método POST. No backend, a função associada a '/products' processa o texto usando a API da OpenAI para identificar e extrair os produtos mencionados. Após a extração inicial, a função realiza *scraping* detalhado dos produtos do site Auchan para obter informações adicionais como preço, links e outros detalhes relevantes. Finalmente, os resultados são apresentados ao utilizador numa nova página (*products.html*), onde ele pode visualizar os produtos extraídos e selecionar os que deseja explorar mais profundamente.

4.1.4 Detalhes Técnicos

- **Flask:** Uma framework de Python para desenvolvimento web.
- **OpenAI API:** Utilizada para processamento de linguagem natural, extração de informações dos textos inseridos e para responder a perguntas do utilizador.

- Bibliotecas Python como **dotenv** para gestão de variáveis de ambiente e **webscraping** para extrair dados de sites web.

4.2 Seleção de produtos

Produtos

Selecione os produtos que deseja ter mais informações.

Produto: pão

Nome	Preço	Selecionar
pão de rio maior 900g	2.19€	<input checked="" type="checkbox"/>
pão de trigo auchan cultivamos o bom cereais do alentejo 800g	2.31€	<input type="checkbox"/>
bolinhas caruço & filhos de mafra 5x70g	1.35€	<input type="checkbox"/>
pães caruço & filhos de mafra fatiados 400g	1.29€	<input type="checkbox"/>

queijo mozzarella auchan fatias 200 g	1.89€	<input type="checkbox"/>
queijo agros flamengo fatias 200g	2.19€	<input type="checkbox"/>
pizza fiambre e queijo auchan 405 g	2.49€	<input type="checkbox"/>
queijo creme philadelphia regular 150gr	1.79€	<input type="checkbox"/>
crujitos auchan sabor a queijo 60g	0.55€	<input type="checkbox"/>

Escolher

Voltar à Página Inicial

Fig. 2 Página para seleção dos produtos

A página `/productsSelection` é parte essencial do fluxo da aplicação Flask, onde os utilizadores selecionam os produtos dos quais desejam obter mais informações detalhadas. Exploreemos em detalhes o conteúdo e a funcionalidade desta página.

4.2.1 Conteúdo da Página

A página é intitulada como “Produtos” e tem um formulário de seleção de produtos utilizado para enviar os produtos selecionados para a rota `/productsSelection` via método `POST`. Cada produto é exibido numa secção separada identificada pelo seu nome e para cada produto é exibida uma tabela com as colunas nome, preço e seleção que é um checkbox para o utilizador poder selecionar os produtos que deseja. Por fim, existem dois botões, um que permite ao utilizador confirmar a seleção dos produtos e outro que lhe permite retornar à página inicial.

4.2.2 Funcionamento

Após o utilizador submeter o texto na página inicial, a função associada a `/products` extrai os produtos e os seus detalhes usando a API da OpenAI e `scraping` do site Auchan. A página `/productsSelection` recebe os dados dos produtos extraídos e os exibe em forma de tabelas, onde o utilizador pode selecionar quais produtos deseja obter mais informações. Quando o utilizador seleciona os produtos e envia o formulário, a rota `/productsSelection` processa os produtos selecionados, extrai informações adicionais se necessário (como detalhes de produto), e os prepara para exibição na próxima etapa.

4.2.3 Detalhes Técnicos

- **Flask:** Framework de Python utilizado para desenvolvimento web.
- **Templates Jinja2:** Utilizados para renderizar dinamicamente o HTML com os dados dos produtos.
- Integração com **API da OpenAI** e **scraping** do site Auchan para obter detalhes dos produtos.

4.3 Pergunta sobre os produtos

Pergunta sobre os Produtos Seleccionados

Escreva sua pergunta sobre os produtos seleccionados:

Quero saber as informações nutricionais dos produtos e como devo os conservar.



Submeter pedido de informação

Voltar à Página Inicial

Fig. 3 Página de inquérito sobre os produtos seleccionados

A página *questions.html* permite aos utilizadores interagirem com o LLM que assiste sobre a seleção de produtos para fazer perguntas ou solicitações específicas sobre os produtos seleccionados. Analisaremos o conteúdo e a funcionalidade detalhadamente:

4.3.1 Conteúdo da Página:

A página tem um formulário utilizado para capturar a pergunta do utilizador sobre os produtos seleccionados. Os dados são enviados com um método *POST* para a rota */submitQuestion*. Além disso, existem dois botões, um que permite ao utilizador enviar a pergunta e os produtos seleccionados para processamento e outro que oferece ao utilizador a opção para retornar à página inicial.

4.3.2 Funcionamento:

A página *questions.html* serve como interface para os utilizadores poderem fazer perguntas específicas sobre os produtos seleccionados anteriormente. Após o utilizador seleccionar produtos na página */productsSelection*, ele é redirecionado para esta página onde pode formular uma pergunta sobre esses produtos. Quando o formulário é submetido, os dados são enviados para a rota */submitQuestion* da aplicação Flask. Essa rota processa a pergunta usando a API da OpenAI (modelo GPT-3.5) para gerar uma resposta adequada com base nos produtos seleccionados e na pergunta formulada.

4.3.3 Detalhes Técnicos:

- **Flask:** Framework de Python utilizado para desenvolvimento web.
- **Templates Jinja2:** Utilizados para renderizar dinamicamente o HTML com os dados dos produtos seleccionados.
- Integração com **API da OpenAI (GPT-3.5)** para processar e gerar respostas com base na pergunta do utilizador.

4.4 Resposta do LLM

Resposta à Pergunta

Pergunta: Quero saber as informações nutricionais dos produtos e como devo os conservar.

Resposta: Claro! Aqui estão as informações nutricionais e instruções de conservação para os produtos selecionados: 1. **Pão de Rio Maior 900g:** ****Informações Nutricionais:**** Por 100g: Lípidos: 0,65g; hidratos de carbono: 36g; proteínas: 4,6g; sal: 0,7g. ****Conservação:**** Conservar em local fresco e seco. 2. **Posta Salmão Auchan Cultivamos o Bom 4/5 kg:** ****Informações Nutricionais:**** Não especificado. ****Conservação:**** Conservar entre 0°C e 2°C. 3. **Queijo Flamengo Auchan à Mesa em Portugal Açores Fatias 500g:** ****Informações Nutricionais:**** Por 100g: Energia: 337kcal; Lípidos: 26g; Proteínas: 25g; Sal: 1,3g. ****Conservação:**** Conservar entre 2°C e 10°C. Se precisar de mais alguma informação, é só pedir!

Produtos Selecionados:

- pão de rio maior 900g,2.19€, <https://www.auchan.pt/produtos-frescos/padaria/pao-fresco-e-broa/pao-de-rio-maior-900g/2120843.html> **{'infos': {'Quantidade Liquida': '0.9 KG', 'Ingredientes/Composição': 'INGREDIENTES: Farinha de TRIGO Tipo 65 (contém GLÚTEN), Água, Levedura e Sal. Conservar em local seco e fresco.', 'Informações Nutricionais': 'Valores nutricionais médios por 100 g: lípidos: 0,65 g, dos quais ácidos gordos saturados: 0,3 g; hidratos de carbono: 36 g, dos quais açúcares: 1,6 g; fibras: 1,3 g; proteínas: 4,6 g; sal: 0,7 g;', 'Conservação': 'Conservar em local fresco e seco.', 'Denominação': 'PÃO DE RIO MAIOR-900 G', 'Nome e Morada': 'Auchan Retail Portugal - Estrada de Paço de Arcos nº 48-A - 2770-129 Paço de Arcos', 'Informação Adicional': 'Confirmar a informação no rótulo do artigo. Devido a possíveis alterações de embalagens e/ou rótulos, deverá considerar sempre a informação que acompanha o produto que recebe.'}}}**
- posta salmão auchan cultivamos o bom 4/5 kg,15.99€, <https://www.auchan.pt/produtos-frescos/peixaria/peixe-fresco/posta-salmao-auchan-cultivamos-o-bom-45-kg/3289851.html> **{'infos': {'Ingredientes/Composição': 'Peixe', 'Conservação': 'Conservar entre 0° e 2°C', 'Informações': 'Os artigos Cultivamos o Bom assentam sobre 4 pilares: Qualidade; Respeito pelo Ambiente; Socialmente Responsável; Economicamente viáveis. Qualidade É um produto com elevados padrões de qualidade que se traduz em maior frescura e sabor. Respeito pelo Ambiente A sua produção é feita de forma sustentável e respeitando toda a biodiversidade. Socialmente Responsável A sua produção apoia as comunidades e pequenos produtores, numa estratégia que visa criar mais empregos, equidades e maior produção local em cada comunidade. Economicamente Viável É pago um preço justo ao produtor de forma a que este obtenha um rendimento sustentável que permita desenvolver a sua produção.', 'Denominação': 'POSTA SALMÃO AUCHAN CULTIVAMOS O BOM 4/5 KG', 'Nome e Morada': 'Mowi Iberia SL', 'Origem': 'Noruega', 'Informação Adicional': 'Confirmar a informação no rótulo do artigo. Devido a possíveis alterações de embalagens e/ou rótulos, deverá considerar sempre a informação que acompanha o produto que recebe.', 'Sugestões Utilização': 'Cozinhar antes de consumir.'}}}**
- queijo flamengo auchan à mesa em Portugal açores fatias 500g,3.99€, <https://www.auchan.pt/produtos-frescos/queijaria/queijo-fatiado-e-barra/queijo-flamengo-auchan-a-mesa-em-portugal-acores-fatias-500g/3352009.html> **{'infos': {'Quantidade Liquida': '0.5 KG', 'Ingredientes/Composição': 'LEITE de vaca pasteurizado, sal, cloreto de cálcio, coalho animal, fermentos lácticos e conservante (lisozima de OVO)', 'Informações Nutricionais': 'Valores nutricionais médios por 100g: Não Preparado: Energia: 1398.00 kJ / 337.00 kcal; Lípidos: 26.00 g. Dos quais ácidos gordos saturados: 18.00 g; Hidratos de carbono: 0.50 g. Dos quais açúcares: 0.50 g; Proteínas: 25.00 g; Sal: 1.30 g', 'Conservação': 'Conservar entre 2°C e 10°C', 'Informações': 'Embalado em atmosfera protetora.', 'Denominação': 'Queijo Flamengo Fatias', 'Preparação': 'Retirar o produto do frigorífico e da embalagem 10 minutos antes de consumir.', 'Nome e Morada': 'Auchan Retail Portugal, S.A., Estrada de Paço de Arcos, n.º 48A 2770129 Paço de Arcos', 'Origem': 'Origem: Portugal Açores.', 'Nutri-score': 'D', 'Queijo em Fatias (LS)': 'Queijo Flamengo'}}}**

[Voltar à Página Inicial](#)
[Fazer mais uma pergunta](#)

Fig. 4 Página com a resposta do LLM

A página ‘answer.html’ na sua aplicação Flask é responsável por exibir a resposta gerada pelo modelo de linguagem GPT-3.5 da OpenAI em resposta à pergunta feita pelo utilizador sobre os produtos selecionados. Analisaremos o conteúdo e a funcionalidade detalhadamente.

4.4.1 Conteúdo da Página

A página exibe a pergunta feita pelo utilizador e a resposta da mesma, gerada pelo modelo de linguagem. Abaixo da resposta é apresentado uma lista com os produtos selecionados pelo utilizador. Por fim, existem dois botões, um que permite ao utilizador retornar à página inicial da aplicação e outro que permite ao utilizador retornar à página que permite fazer uma nova pergunta.

4.4.2 Funcionamento

A página *answer.html* permite apresentar de forma clara e organizada a pergunta feita pelo utilizador, a resposta gerada pelo modelo de linguagem e a lista de produtos sobre os quais a pergunta foi feita. Após o utilizador enviar uma pergunta na página *questions.html*, os dados são processados pela rota */submitQuestion*. A função associada a */submitQuestion* utiliza a API da OpenAI para gerar uma resposta baseada na pergunta e nos produtos selecionados, sendo então exibida nesta página (*answer.html*), com a pergunta e a lista de produtos selecionados. Por fim, o utilizador tem a opção de retornar à página inicial ou fazer outra pergunta sobre os mesmos produtos.

4.4.3 Detalhes Técnicos

- **Flask:** Framework de Python utilizado para desenvolvimento web.
- **Templates Jinja2:** Utilizados para renderizar dinamicamente o HTML com os dados da pergunta, resposta e produtos selecionados.
- **Integração com API da OpenAI (GPT-3.5)** para processar e gerar respostas com base na pergunta do utilizador.

5 Conclusões e Trabalho Futuro

O sistema desenvolvido demonstrou ser uma ferramenta viável e eficaz para a seleção automatizada de produtos alimentares e bebidas, utilizando técnicas avançadas de PLN e *scraping* de dados. O impacto potencial desta tecnologia no setor de venda de produtos alimentares é significativo, proporcionando uma vantagem competitiva ao facilitar decisões informadas e rápidas.

Para o futuro, planeia-se expandir o sistema para integrar mais fontes de dados e otimizar algoritmos de *scraping* para lidar com mudanças nas estruturas das páginas web. Além disso, pretende-se explorar melhorias na interface do utilizador para tornar o sistema mais intuitivo e acessível aos utilizadores finais, incorporando feedbacks recebidos durante os testes iniciais.